DOCUMENT RESUME

ED 036 873 24 CG 005 172

AUTHOR Postman, Leo
TITLE Comments on Professor Gagne's Paper Entitled "Instructional Variables and Learning Outcomes." Center for the Study of Evaluation of Instructional Programs Occasional Report No. 18.
INSTITUTION California Univ., Los Angeles. Center for the Study of Evaluation.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-6-1646
PUB DATE Sep 68
CONTRACT OEC-4-6-061646-1909
NOTE 15p.; Paper from the Proceedings of the Symposium on Problems in the Evaluation of Instruction, Los Angeles, California, December 13--15, 1967

EDRS PRICE EDRS Price MF-$0.25 HC-$0.85
DESCRIPTORS *Cognitive Measurement, *Educational Objectives, Educational Philosophy, Evaluation Techniques, Instruction, Learning, *Measurement, *Measurement Goals, *Objectives

ABSTRACT

Professor Postman is in agreement with Professor Gagne's view that valid measurement of learning outcomes is an essential part of the evaluation of educational systems. The paper stresses that the categorization of outcomes and measuring operations should be regarded as flexible heuristic devices, and that it is desirable to guard against the risk of standardization. It also discusses the classes of outcomes distinguished in experimental investigation, and focuses on the selection of operations which will both avoid distortion of measurement and distinguish the outcomes under study from the action of other variables. (Author)

Occasional Report No. 18

COMMENTS ON PROFESSOR GAGNÉ'S PAPER ENTITLED "INSTRUCTIONAL VARIABLES AND LEARNING OUTCOMES"

Leo Postman

Center FOR THE
Study of Evaluation
OF INSTRUCTIONAL PROGRAMS

University of California, Los Angeles, September 1968

CO-DIRECTORS

Merlin C. Wittrock Erick L. Lindman

ASSOCIATE DIRECTORS

Marvin C. Alkin Frank Massey, Jr. C. Robert Pace

The CENTER FOR THE STUDY OF EVALUATION OF INSTRUCTIONAL PROGRAMS is engaged in research that will yield new ideas and new tools capable of analyzing and evaluating instruction. Staff members are creating new ways to evaluate content of curricula, methods of teaching and the multiple effects of both on students. The CENTER is unique because of its access to Southern California's elementary, secondary and higher schools of diverse socio-economic levels and cultural backgrounds.

COMMENTS ON PROFESSOR GAGNE'S PAPER ENTITLED "INSTRUCTIONAL VARIABLES AND LEARNING OUTCOMES"

Leo Postman

University of California, Berkeley

From the Proceedings of the

SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

University of California, Los Angeles
December, 1967

M. C. Wittrock, Chairman

Sponsored by the Center for the
Study of Evaluation of Instructional Programs

The research and development reported herein was performed pursuant to a contract with the United States Department of Health, Education, and Welfare, Office of Education under the provisions of the Cooperative Research Program.

CSEIP Occasional Report No. 18, September 1968
University of California, Los Angeles

COMMENTS ON PROFESSOR GAGNÉ'S PAPER ENTITLED "INSTRUCTIONAL VARIABLES AND LEARNING OUTCOMES"

Leo Postman

Professor Gagné has presented to us a very useful conceptual framework for the assessment of learning outcomes. I find myself in full agreement with his view that the valid measurement of learning outcomes is an essential and, indeed, a critical part of the evaluation of educational systems. I want to underscore especially his call for the use of measures of retention and of transfer in the assessment of learning outcomes. Since the objective of instruction is preparation for future activities, long-term retention and the potentialities for transfer to the mastery of new tasks must be the primary criteria of the success of our methods of teaching.

A major value of Professor Gagné's contribution lies in the combination of a general analysis of principles of measurement with the development of an orderly system for the classification of learning outcomes. There is likely to be wide agreement on his formulation of the criteria which must be met by valid measures of learning outcomes, namely, distinctiveness and freedom from distortion. It is, indeed, essential to differentiate as precisely as possible the measuring operations for various classes of learning outcomes and to minimize the influence of uncontrolled and irrelevant variables on performance.

We would expect less consensus when we consider the translation of these general principles into specific operations. The

choice of categories of outcomes and of defining operations is based both on theoretical presuppositions and on generalizations from available empirical findings. There will be legitimate differences of opinion regarding the most useful categorization of learning outcomes and the empirical generalizations supported by the experimental evidence. The discussion of such differences in opinion should in the long run contribute to the improvement of our methods of assessment. With this objective in mind, I should now like to comment on some of Professor Gagné's proposals for the application of the basic criteria of measurement to the systematic assessment of learning outcomes. My purpose will be not so much to disagree with the substance of his recommendations as to emphasize the close interrelations between theory, inferences from available empirical data, and the establishment of criteria for new measurements. The burden of my argument will be that the categorization of outcomes and the measuring operations coordinated with them should be regarded as flexible heuristic devices, and that it will be well to guard against the risk of standardization. I believe that is also Professor Gagné's intention, but it may be useful to focus explicitly on this point.

Let us consider first the measurement of the three broad classes of outcomes distinguished traditionally in our experimental investigations: learning, retention, and transfer. To measure each of these outcomes distinctively requires the performance of a control operation to support the conclusion that the observed outcome reflects the process under study rather than some other process. Once this requirement is accepted, the choice of the

appropriate control operations is not necessarily self-evident and often must be based on theoretical decisions. Take the distinctive measurement of transfer as an example. As Professor Gagné formulates the problem, "the inference one wishes to make is that some capability a ... has an effect on the learning of capability b. The capabilities a and b are different in some respects that are specifiable." To insure distinctiveness, a control group is used to demonstrate that without prior acquisition of a there is no effect on the learning b. And the control operations must be such as to permit the inference that "capability a has transferred, not simply that capability b has been facilitated or interfered with."

These defining operations are clearly appropriate to the measurement of specific transfer, i.e., transfer based on the similarity relations between elements in the successive tasks. The capabilities a and b are different in specifiable ways. The transfer of a influences the learning of b because there is some systematic relation between these different capabilities. Facilitation of b or interference with b is not included within the domain of transfer. The classical paradigms of specific transfer, such as the learning of old responses to new stimuli, can be readily subjected to tests of distinctiveness implied by these boundary conditions. However, it would be difficult to apply these particular criteria of distinctive measurement to another type of transfer which contributes heavily to the progressive improvement of the learner in the performance of successive tasks, viz., general transfer or learning to learn.

In studies of learning to learn, the experimental procedures are explicitly designed to minimize the possibility that the observed improvements reflect the transfer of specific components from earlier tasks; thus, every effort is made to eliminate overlap of elements or similarities between the units in the successive tasks. As the individual goes from task A to task B, the corresponding capabilities a and b are assumed to be instances of the same class with reference to general transfer or learning to learn. The question at issue is whether the individual changes his mode of attack on B as a result of his experience with A; in that sense transfer is the facilitation of b rather than the carry-over of a specific a. It is our hope, of course, to be able through experimental analysis to break down general transfer into component habits and skills. However, in embarking on studies of learning to learn we were unable to specify these component habits and skills in advance. Rather, the component sources of transfer must be induced gradually from the conditions and characteristics of improvement and the validity of the inductions when tested experimentally. Thus, as our knowledge of general transfer increases, the experimental paradigms used in its analysis progressively approach those of specific transfer.

The basic point here is that the control operations designed to demonstrate distinctiveness will be different for specific transfer and for general transfer. In fact, it is now standard procedure in experiments on specific transfer to include control operations to parcel out the effects of general transfer. Thus, I am in no way questioning the defining operations for transfer

proposed by Professor Gagné but am merely trying to delimit the domain of outcomes to which they apply. It is important that our approach to the problem of transfer be in no way constrained by any existing set of criteria of distinctiveness. The same applies, of course, to other learning outcomes. It is a truism worth emphasizing that theoretical considerations must take precedence over the constraints imposed by conventional operational definitions.

While the criterion of distinctiveness must be recognized as essential for orderly measurement, we must also guard against the possibility that a compartmentalization of outcomes unduly restricts our inferences about the effects of training. It is possible to point to situations in which outcome B provides the best, and perhaps the only means for establishing the existence of outcome A because the usual criteria for A lack sensitivity. Consider the situation in which an individual is given several practice trials on a task, say on a list of paired associates. An examination of his performance on these practice trials shows that there are a number of items which he never gave correctly, and he is also unable to recall any of these items correctly on a test of retention immediately after the end of practice. By the usual distinctive criteria, he has failed to learn these items. There is clear experimental evidence, however, that under these circumstances a test of transfer will show that some degree of learning for these items did occur on the practice trials. For example, if the stimuli in these items are paired with new responses, there will be significant negative transfer relative to an appropriate control

baseline. The degree of learning on the practice trials was insufficient for the performance of the prescribed task, but it was great enough to result in a significant amount of transfer. It may be mentioned in passing that the use of a transfer test for determining whether or not learning has occurred has become an accepted procedure in experimental evaluations of the hypothesis that association is an all-or-none process.

More generally, a transfer test must typically be used to decide between alternative interpretations of what an individual has learned in a given training situation. The analysis of stimulus selection is a case in point. Suppose the learning task consists of paired associates in which the stimulus terms are nonsense syllables presented against colored backgrounds, these colors being clearly different for each syllable. Once the subject has mastered the list, the question arises of what stimuli he has learned to respond to--the syllables, the colors, or both. That is, did he select one of the components of the stimulus terms and associate his responses to that component? The answer to the question of what he has learned depends on whether and to what extent there has been such stimulus selection. And the answer can be obtained only by a test of transfer, i.e., by determining the subject's ability to give the correct responses to each of the components separately and in combination. Thus, the logic of the experimental question makes it necessary to draw inferences about learning by means of the distinctive operations of transfer. It is always necessary to insist that the operations of measurement be distinctive, but the acceptance of the criterion of

distinctiveness does not entail an invariant identification of a given type of outcome with a particular set of measuring operations.

When the criterion of distinctiveness is applied to measures of outcomes of varying complexity, as is done in the hierarchical scheme proposed by Professor Gagné, it is important not to assume on a priori grounds that the more complex outcomes represent a compounding of the less complex ones. Suppose that a response chain has been established and that it is possible to specify and to measure independently and distinctively each of the links in the chain. We must be cautious about making the inference that the integrated chain represents a summation of these independent elements, and the same holds for any capability which can be analyzed into distinctive and independently measurable components. The two-stage conception of paired-associate learning to which Professor Gagné refers is a case in point. The distinction between response learning and associative learning has proved very useful in the analysis of the conditions and characteristics of paired-associate learning. Each component can be measured independently, for example, by free recall of the responses on the one hand and a test of stimulus-response matching on the other. But it is also a fact that at a given stage of practice the subject's performance in the paired-associate task does not represent a simple summation of the two components measured independently. When a high-order capability is critically dependent on the interaction of conceptually distinct components, two-stage measurement yields only limited information about the status of

such a complex capability. The need then arises for the development of distinctive measures directed at the assessment of the interaction of the components.

The decisive influence of theory on the formulation of the criteria of valid measurement is exhibited most clearly in relation to the requirement of freedom from distortion. To avoid distortion of measurement we must choose a set of operations which distinguishes the outcome under study from the action of other variables. To apply this principle we must then distinguish between relevant and irrelevant variables, and that is clearly a theoretical decision. Such decisions will obviously be subject to disagreement on theoretical grounds. I will limit myself to one illustrative example where I found myself in disagreement with Professor Gagné's specification of potential sources of distortion. In describing the measuring operations for the assessment of retention, he states that distortion is prevented by operations which control, among other things, the kind of intervening activity known to produce varying amounts of interference, i.e., retroactive inhibition. But according to at least one theoretical view, forgetting is always the result of interference, whether this interference is produced by formal laboratory tasks or by the activities of the learner outside the laboratory. On this theoretical assumption it becomes impossible in principle to eliminate, or to hold strictly constant, interference from intervening activities in the measurement of long-term retention. One can try to influence the probability of interference by choosing materials assumed to have certain similarity relations to the learner's outside activities.

Clearly, an interference theorist, a decay theorist, and a consolidation theorist will have quite different conceptions of what are distortion-free conditions for the measurement of retention. Given the lack of theoretical agreement among experimental investigators, the same would hold true for most if not all outcomes of learning.

As I indicated at the outset, I find myself in substantial agreement with Professor Gagné's basic approach to the measurement of learning outcomes. In dissenting from some of his proposals, it has been my intention to bring to the fore the inevitable theoretical basis of decisions about measurement. A good theory will generate appropriate methods of measurement, but the continued acceptance of currently successful measuring operations will not necessarily lead to better theories.

PAPERS FROM THE SYMPOSIUM ON PROBLEMS IN THE EVALUATION OF INSTRUCTION

M. C. Wittrock, Chairman

THEORY OF EVALUATION OF INSTRUCTION

Benjamin Bloom, Principal Speaker University of Chicago	*Toward a Theory of Testing Which Includes Measurement-Evaluation-Assessment*, Occasional Report No. 9
Michael Scriven, Discussant University of California, Berkeley	*Comments on Professor Bloom's Paper*, Occasional Report No. 10
Gene Glass, Discussant University of Colorado	*Comments on Professor Bloom's Paper*, Occasional Report No. 11
J. P. Guilford, Discussant University of Southern California	*Comments on Professor Bloom's Paper*, Occasional Report No. 12
Robert Glaser, Principal Speaker University of Pittsburg	*Theory of Evaluation of Instruction*, Occasional Report No. 13
Robert Stake, Discussant University of Illinois	*Comments on Professor Glaser's Paper*, Occasional Report No. 14
Arthur Lumsdaine, Discussant University of Washington	*Comments on Professor Glaser's Paper*, Occasional Report No. 15

INSTRUCTIONAL VARIABLES

Robert Gagné, Principal Speaker University of California, Berkeley	*Instructional Variables and Learning Outcomes*, Occasional Report No. 16
Richard Anderson, Discussant University of Illinois	*Comments on Professor Gagné's Paper*, Occasional Report No. 17
Leo Postman, Discussant University of California, Berkeley	*Comments on Professor Gagné's Paper*, Occasional Report No. 18

CONTEXTUAL VARIABLES

Dan Lortie, Principal Speaker University of Chicago	*The Cracked Cake of Educational Customs and Emerging Issues in Evaluation*, Occasional Report No. 19
C. Wayne Gordon, Discussant University of California, Los Angeles	*Comments on Professor Lortie's Paper*, Occasional Report No. 20
N. L. Gage, Discussant Stanford University	*Comments on Professor Lortie's Paper*, Occasional Report No. 21

CRITERION VARIABLES

Samuel Messick, Principal Speaker Educational Testing Service	*The Criterion Problem in the Evaluation of Instruction: Assessing Possible Not Just Probable Intended Outcomes*, Occasional Report No. 22
Paul Blommers, Discussant University of Iowa	*Comments on Dr. Messick's Paper*, Occasional Report No. 23
Leonard Cahen, Discussant Educational Testing Service	*Comments on Dr. Messick's Paper*, Occasional Report No. 24
Marvin Alkin, Principal Speaker University of California, Los Angeles	*Evaluating Cost-effectiveness of Instructional Programs*, Occasional Report No. 25
Marvin Hoffenberg, Discussant University of California, Los Angeles	*Comments on Professor Alkin's Paper*, Occasional Report No. 26
John Bormuth, Discussant University of Chicago	*Comments on Professor Alkin's Paper*, Occasional Report No. 27

METHODOLOGICAL ISSUES

David Wiley, Principal Speaker University of Chicago	*Design and Analysis of Evaluation Studies*, Occasional Report No. 28
Chester Harris, Discussant University of Wisconsin	*Comments on Professor Wiley's Paper*, Occasional Report No. 29

Theodore Husek, Discussant
University of California,
Los Angeles

Comments on Professor Wiley's Paper, Occasional Report No. 30

Martin Trow, Principal Speaker
University of California,
Berkeley

Methodological Problems in the Evaluation of Innovation, Occasional Report No. 31

Eugene Litwak, Discussant
University of Michigan

Comments on Professor Trow's Paper, Occasional Report No. 32

David Nasatir, Discussant
University of California,
Berkeley

Comments on Professor Trow's Paper, Occasional Report No. 33